VALIDATION OF THE CONSISTENT-YEAR V003 MODIS LAND COVER PRODUCT

ABSTRACT

We estimate the accuracy of the IGBP layer of our Consistent-Year Land Cover product (V003) to be 75–80 percent globally; 70–85 percent by continental regions; and from 60–90 percent individual classes. These estimates are supported by quantitative analysis of (1) classification of unseen training sites; and (2) confidence values aggregated by land cover class and continental region. Accordingly, we deem the product to be **VALIDATED—LEVEL 1**.

INTRODUCTION

The purpose of the MODIS land cover product validation activity is to provide information to users of our land cover data describing the accuracy of our classifications and their inherent error structure. We use two primary tools to assess the quality of our product: confusion matrixes and aggregations of confidence values. The confusion matrixes describe how well the training sites are classified when they are "unseen" by the classifier, and so provide information on the accuracy of the classification process as applied to the training site database. The confidence values are generated by the classifier and indicate how well the pattern of spectral and temporal variation in annual observations of each pixel fits the examples of training data provided to the classifier. They may be treated as probabilities of correct classification, given the input training data. These tools are fully repeatable, and will be used to assess future releases of the land cover product.

Before we discuss these two analytical tools and their application to validation of the MODIS consistent-year land cover product, it will be helpful to provide an overview of the classification process by which the consistent-year land cover map was made.

THE CLASSIFICATION PROCESS

The Decision Tree Classifier

In outline, the MODIS land cover product uses a supervised classification approach in which training sites are provided to a decision-tree classifier. The classifier then generates a decision tree that is exercised on the global data field, thus making a global map. This process is described in more detail in Friedl et al. (2002).

The classification is enhanced by *boosting*. In this process, training data are input to the classification algorithm and a decision tree is estimated. The decision tree is then exercised on the training data. The outcome is compared to the input training data and a new decision tree is estimated, but this time the training pixels that are incorrectly classified are weighted more heavily. This new decision tree is in turn exercised, and based on its output, the training sites are again reweighted. The process continues until 10 boosted decision trees are prepared. The final classification of each pixel uses all 10 trees, which are taken as experts that vote on the proper label for the pixel. The label is then assigned using a plurality rule. Statistical theory shows that the voting process using boosted trees can be used to estimate the probability that a pixel belongs to each class. This allows direct application of prior probabilities to the classification output, a technique that provides a powerful tool for adding prior knowledge to improve the global classification map (McIver and Friedl, 2002).

Prior Probabilities

Given a spectral signature for a pixel and a set of training sites, the boosted classification process yields a set of probabilities of membership of the pixel in each of the classes. The class label can then be assigned using the class associated with the highest probability. However, the accuracy of a final map can be improved by using prior probabilities. *Prior probabilities* specify how likely each class is to appear at each geographic location, based on prior knowledge. By using Bayes' rule to combine the classification probabilities with the prior probabilities for each pixel, a set of *posterior probabilities* is calculated that merges the two sets. The class label is then assigned using the posterior, rather than the classification, probabilities.

For example, suppose a dark pixel from a lava flow in the Sahara Desert is classified as water. By specifying that the prior probability of open water in the Sahara is low but the probability of a barren or sparse land cover is high, the posterior probability for water decreases and the posterior probability for barren/sparse increases. This change shifts the label assigned from water to barren/sparse.

It is also possible to use multiple sets of prior probabilities in succession to combine different types of prior information. The MODIS consistent-year land cover product includes the application of four sets of prior probabilities. These are:

1. Priors from the IGBP DISCover I land cover map (MODIS at-launch land cover product). These priors were determined for each pixel by passing a 200 by 200 pixel moving window across the IGBP DISCover I land cover map and using the proportions of land covers within that window as estimates of prior probabilities. This amounts to a smoothing operation that tends to preserve the general distribution patterns of land cover types from the IGBP DISCover heritage product. In combining classification probabilities with this set of priors, the priors were weighted at 50 percent.

2. Priors from the Foley map of global agricultural intensity. This map provides an index of agricultural activity at a 1/2° spatial resolution. We smoothed the coarse grid cell structure of this map to 1 km by using a 200 by 200 moving window, then applied a prior-probability adjustment to cropland and noncropland classes. This prior source was also weighted at 50 percent.

3. Correction for sample distribution. The decision-tree classifier optimizes the classification for the training samples at hand. If one class dominates the training sites, the classifier will respond by creating more nodes for that class. If another class is represented by only a few samples, the classifier creates fewer nodes. This property tends to integrate the class distribution of the training sites into the structure of the classifier. To remove this effect, we reverse-weight the classification probabilities by the distribution of training site pixels. The reverse weighting lowers the classification probabilities of the classes that are better represented in the training sample and raises the probabilities of those that are less well represented in the sample.

4. Global land cover class distribution. As a last correction, the probabilities are weighted by a single set of prior probabilities—the entire global distribution of land cover classes from the IGBP DISCover map. This preserves the general balance of land cover types across all continents and regions.

CONFUSION MATRIX

The *confusion matrix* is a commonly used tool for assessment of accuracy for land cover classifications. The matrix scores how the classification process has labeled a series of test sites or test pixels at which the correct land cover label is known. Typically, the true class label is displayed across rows, while the actual mapped class is displayed in columns. The diagonal of the confusion matrix displays the number of sites or pixels for which the true class and the mapped class agree. The *overall accuracy* of the entire sample is then the sum of the diagonal elements divided by the total of all sites or pixels. For individual classes, the marginal totals of the matrix can easily be used to estimate the producer's accuracy and user's accuracy from the sample. The *producer's accuracy* is the probability that a pixel truly belonging to class i is also mapped as class i, while the *user's accuracy* is the probability that a pixel mapped as class i is truly of class i. Using marginal and diagonal totals to estimate these accuracies, however, is subject to bias if the proportions of classified sample sites or pixels across classes is different from the proportions of classes in the output map. To remove these biases, the proportions of classes observed for the entire map are used in computation (Card, 1982).

Note that for the confusion matrixes presented here, the classifier label is assigned after the application of prior probabilities as described above. In this way, the output classification used in validation is most similar to the output used to construct the map. If output class labels are compared to true classes before application of priors, the confusion matrix scores fewer errors and yields higher producer's and user's accuracies. Thus, the prior probabilities have the effect of detuning the classifier slightly and producing a smoother map that, in our judgment, is a better output.

Sample Design

Proper statistical characterization of accuracy as measured by a confusion matrix depends on a proper sample design for choosing test sites or pixels. Typical sample designs are random, random stratified, and random systematic. These have different implications for determining both overall accuracy and the accuracy of individual classes.

In a random sample, each pixel (assuming equal-area pixels) has the same probability of being chosen as all other pixels. The random sample is the simplest design and is the most efficient for determining the overall accuracy. However, it has the drawback that small or rare land cover classes may be rarely sampled, if at all, and thus the accuracies of individual classes are not known with the same level of precision. That is, large classes will tend to have more samples, while smaller classes will have fewer.

The solution to this problem is the random stratified sample, in which a fixed number of test sites or pixels is drawn from each mapped class. Thus, the smallest and largest classes are sampled using the same number of samples, and the confidence intervals placed on within-class accuracies are comparable. The random stratified sample requires that all sites or pixels within a mapped class have an equal probability of being sampled. Thus, each pixel of a small or rare class is more likely to be sampled than a pixel of a large class. But since we know the area of each class, we can still find a proper overall accuracy estimate by weighting the accuracy of each class by its area.

The random systematic design overcomes the problem that a chance throw of samples onto a geographic grid may leave large portions of the grid sparsely sampled or even entirely unsampled. Since land cover has a broad geographical component based on climate, ecology, and human land use patterns, it is important to ensure that there is substantial representation in the sample from all important regions. Thus, a global map may be sampled by continents, with equal or at least fixed sample sizes allocated to each continental region. As in the case of the random stratified sample, the area of each continent is known and is used to weight the overall accuracy calculation. For smaller regions, the area may be divided into a coarse equal-area grid, with equal numbers of samples drawn from each grid cell. Note also that a sample may be both stratified and systematic. This is probably the best overall type of sample for validation of a global map product, provided that the sample size is large enough to capture the variance of rarer classes with acceptable precision.

The major drawback of any random sample design as applied to a global map product is cost. For example, the true land cover class for a site or pixel is best determined by visiting the location on the ground. Clearly, this is not practical for a global random sample. As a result, pathfinder efforts in global land cover validation have used high-resolution satellite imagery, such as Landsat data, to "visit" sample sites and determine their proper land cover class (Scepan, 1999). Even with this approach, however, the cost of acquiring recent and useful imagery and carrying out its proper interpretation is typically prohibitive.

MODIS Land Cover Approach

Since we lack sufficient resources to conduct proper random sampling, we instead look to the classification of the samples we have on hand—*i.e.*, the training sites. If the training sites are truly representative of the range of variation in each land cover type, then our training sample may also serve as a validation sample. In this approach, the true label of the training site and its pixels is compared to the labels assigned to the pixels by the classifier. However, the accuracies reported will be biased toward high values, since the classifier has already seen the data used for testing. To avoid this problem, we adopted the following procedure:

1. Randomly divide the population of train/test sites into 10 groups of nearly equal size.
2. For each group, train the decision-tree classifier using the remaining 9 groups of sites, thus holding the group unseen from the classification process.
3. Classify the pixels within the sites of the group and report the result to the confusion matrix.

Note that the division is made by random assignment of sites, not pixels. The reason is that pixels at an individual training site tend to be correlated, so that only a few training pixels from each site are needed to convey the signal of that site to the classifier. In a random assignment of pixels in which the classifier sees 90 percent of the pixels, the classifier will actually get a look at nearly all the training sites. The result will be an inflated accuracy that does not truly reflect the ability of the classifier to generalize beyond the sites it has already seen. In our trials, random sampling of pixels typically produces accuracies 10–15 percent higher than random sampling of sites. We think the lower values are more likely to represent the true accuracy of the MODIS product.

Note also that our final map is made using all training sites. Because all the information is used in producing the final map, we think it is likely to have a slightly higher accuracy than that estimated by using ten, slightly different, 90-percent sample decision trees.

Continental Regions

To investigate the broad range of accuracy values across the continents, we determined accuracies within continental regions as well as globally. Our five continental regions include North America, South America, Africa, Eurasia, and Australia-Insular Asia. (Although there is some small land area in Antarctica that is not snow or ice covered, it is not included in the consistent-year land cover product.) These five regions are bounded by whole 10-degree tiles in the MODIS sinusoidal Level 3 grid as shown in Figure 1. Note that some tiles are considered to be part of two regions. They are identified as “Common Areas” in the graphic. For example, the four tiles centered on the Mediterranean Sea are included in both Eurasia and Africa in the generation of continental statistics.

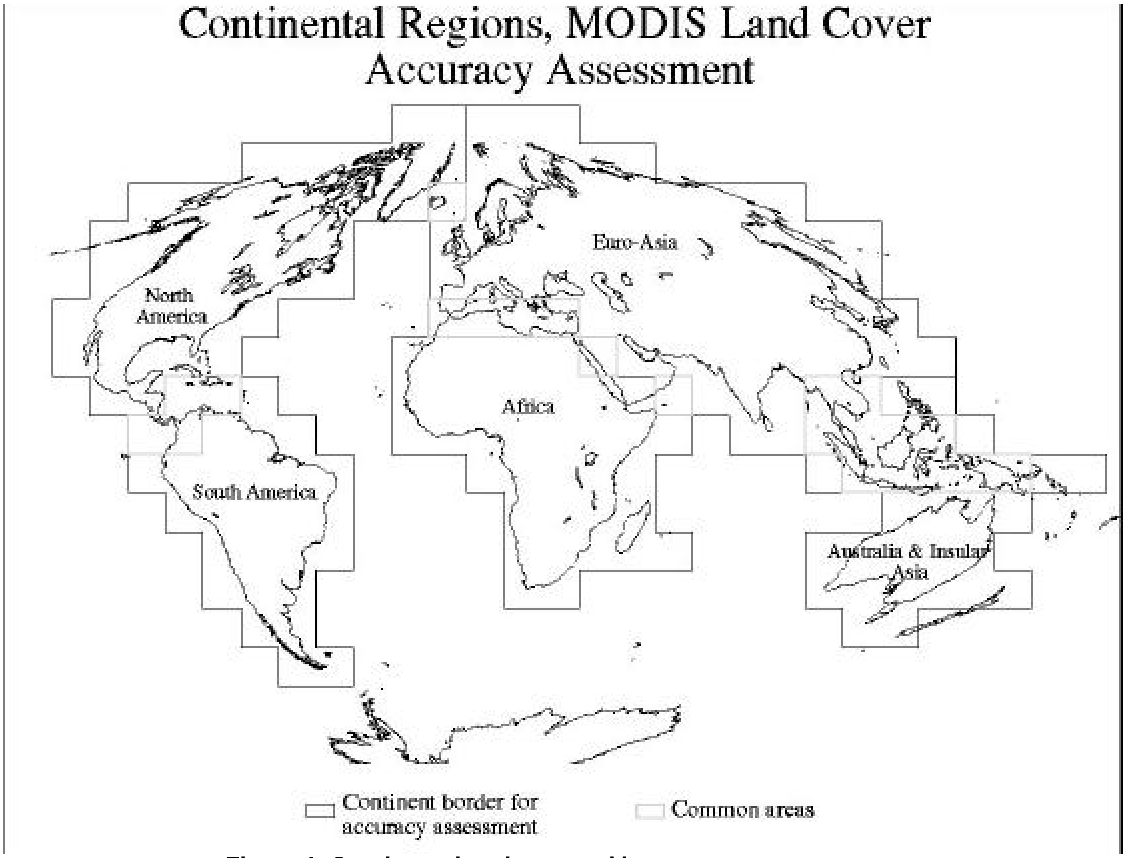

Figure 1. Continental regions used in accuracy assessment.

Training Site Distribution

Table 1 presents the distribution of sites and pixels by class. A total of 1,370 training sites including 39,472 pixels are used in this analysis. The number of sites varies significantly by land cover type, with cropland accounting for the largest number of training sites, followed by evergreen broadleaf forest, evergreen needleleaf forest, and barren/sparse sites. Among the classes represented by relatively few sites are permanent wetlands, deciduous needleleaf forest, and closed shrublands. For these types, it is difficult to find homogeneous sites of sufficient size for our application. Snow and ice and water are also represented by small numbers of sites, but as shown by their pixel counts, these sites are much larger.

Note also that urban and built-up lands (class 13) are omitted. In the consistent-year product, the urban class was determined from the Digital Chart of the World. Classification of urban and built-up lands from training sites is particularly problematic with spectral-temporal data alone, since the many components of urban land covers (soil, pavement, vegetation, building materials) do not provide a consistent signal that can be classified with good accuracy.

IGBP Land Cover Class	Training Site Count	Training Pixel Count	Global Pixels Classified	Global Areal Percentage
1. Evergreen Needleleaf	131	2,056	7,100,847	3.92
2. Evergreen Broadleaf	204	5,409	17,583,346	9.72
3. Deciduous Needleleaf	15	261	2,374,908	1.31
4. Deciduous Broadleaf	57	758	2,016,765	1.11
5. Mixed Forest	96	2,077	8,209,766	4.54
6. Closed Shrubland	20	466	1,068,970	0.59
7. Open Shrubland	87	1,679	31,929,221	17.75
8. Woody Savanna	55	1,167	10,702,581	5.92
9. Savanna	44	1,098	11,218,832	6.20
10. Grasslands	87	1,474	12,363,432	6.83
11. Permanent Wetlands	13	289	559,675	0.31
12. Cropland	263	6,240	17,087,489	9.44
14. Cropland/Nat Veg Mosaic	72	1,447	5,660,478	3.13
15. Snow and Ice	10	1,346	16,501,715	9.12
16. Barren/Sparse	108	4,492	21,977,613	12.15
17. Water	63	9,213	14,575,749	8.06
Total	1,370	39,472	180,928,968	100.00

Table 1. Global counts of sites and pixels by land cover class.

Table 2. presents the distribution of training sites and pixels by continental region. (Because of overlap among regions, site and pixel counts will total larger than the global counts shown in Table 1.) The Eurasian continental region, by virtue of its large size, shows the largest number of sites. North and South America are about equally represented in sites, although the South American sites tend to be smaller. African test sites are significantly fewer. The Australia and Insular Asia region has the fewest sites, which is partly due to the large Australian desert and the vast areas of broadleaf evergreen forests in equatorial Asia. Figure 2 plots the locations of the training sites.

Table 2. Site and pixel counts by region.				
IGBP Land Cover Class	**Training Site Count**	**Training Pixel Count**	**Global Pixels Classified**	**Global Areal Percentage**
Global	1,370	39,472	180,928,968	100.0
North America	368	13,731	30,918,663	17.1
South America	321	8,030	22,181,052	12.3
Eurasia	560	13,290	71,275,640	39.4
Africa	194	5,744	38,711,576	21.4
Australia-Insular Asia	46	1,766	18,046,575	10.0

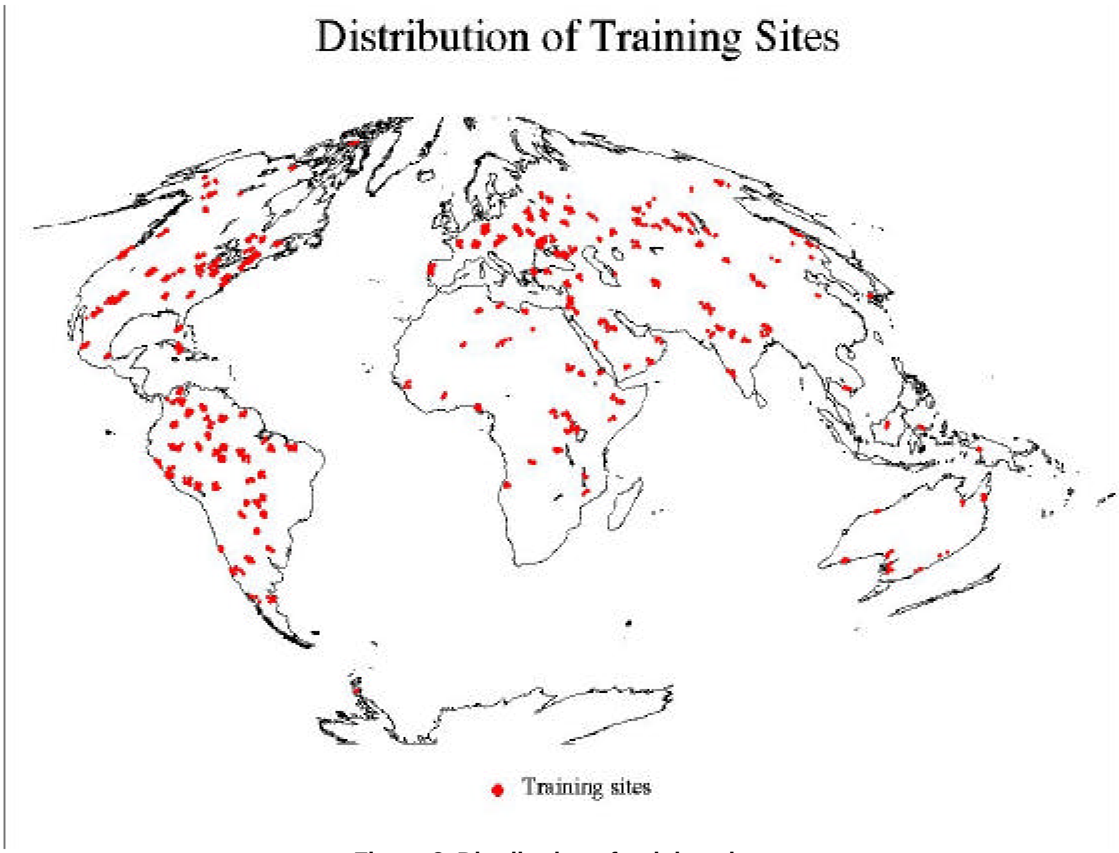

Figure 2. Distribution of training sites.

Global Confusion Matrix

Table 3 presents the global confusion matrix for unseen test sites. The data show the counts of pixels tabulated by training site class and output class label. The table counts nearly 40,000 pixels from 1,370 training sites.

Table 3. Confusion matrix based on classification of unseen test sites (pixel counts).

Training Site Label	Output Class Label: 1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	14.	15.	16.	17.	Row Total
1. Evrgrn Ndlleaf	1323	13	65	23	407	7	35	96	11	6	20	35	7	0	2	6	2056
2. Evrgrn Brdleaf	12	5139	0	3	3	2	7	141	48	14	1	18	14	0	3	4	5409
3. Decid Ndlleaf	20	0	102	3	85	0	5	38	1	3	1	3	0	0	0	0	261
4. Dedic Brdleaf	7	11	15	381	243	1	10	34	10	9	0	16	11	0	2	8	758
5. Mixed Forest	167	3	50	178	1370	1	9	59	7	29	70	71	52	0	0	11	2077
6. Closed Shrblnds	24	18	0	0	6	129	154	37	55	14	0	29	0	0	0	0	466
7. Open Shrblnds	4	4	2	17	9	53	1204	27	9	170	3	5	0	1	168	3	1679
8. Woody Savanna	76	56	0	6	61	3	97	617	154	47	0	36	12	0	0	2	1167
9. Savanna	1	53	3	0	4	25	84	303	504	49	7	13	49	0	3	0	1098
10. Grass-land	5	36	0	1	4	1	161	15	69	1028	0	78	20	0	54	2	1474
11. Wetland	60	15	0	1	7	0	9	9	2	8	174	3	1	0	0	0	289
12. Cropland	23	46	3	33	21	15	243	142	252	365	0	4775	299	0	13	10	6240
14. Crp–Vegn Mosaic	2	134	0	195	62	3	9	113	150	29	0	197	546	0	3	4	1447
15. Snow – Ice	1	0	0	0	0	0	31	0	0	3	0	2	0	1261	47	1	1346
16. Barren	2	6	0	2	12	38	491	10	10	56	0	9	2	0	3853	1	4492
17. Water	7	5	0	9	11	1	2	0	2	6	0	12	3	0	0	9155	9213
Column Total	1745	5541	241	879	2334	279	2574	1647	1289	1858	278	5464	1021	1262	4151	9210	39773

Global and Continental Accuracies

While the confusion matrix shows the exact distribution of training and output class labels, it must be further processed to give useful statistics. These are calculated using the confusion table and the proportions of classes within the entire consistent-year product (Table 1) following the theory and examples shown in Card (1982) for random stratified sampling. Table 4 presents results globally and by continental region. Table 5 shows global per-class accuracies and proportions. (Confidence intervals assume variances are asymptotically distributed as normal distributions.)

Table 4. Global accuracy and accuracy of continental regions (percent).				
Region	Accuracy Estimate	Standard Error	95% Confidence Interval	
			Low	High
Global	71.6	0.25	71.1	72.1
Africa	61.7	0.66	60.3	63.0
Austr & Insular Asia	71.9	2.93	66.1	77.8
Eurasia	67.8	0.40	67.0	68.6
North America	61.3	0.62	60.0	62.5
South America	75.4	0.46	74.4	76.3

As shown in Table 4, the global accuracy as estimated by the training site confusion matrix is 71.6 ± 0.25 percent, giving a confidence interval of 71.1, 72.1 percent. The small standard error of 0.25 percent is due to the very large number of samples. Accuracy varies among continental regions from a low of 61.3 percent for North America to a high of 75.4 percent for South America.

Accuracies by Land Cover Class

Table 5 documents per-class accuracies derived from the training site confusion matrix. Producer's accuracy is the probability that true pixels are correctly classified and thus includes only errors of omission. User's accuracy is the probability that mapped pixel labels are correct and thus includes only errors of commission. The areal proportion estimates take the confusion matrix and the mapped class proportions into account, and so are somewhat different from the raw proportions observed (Table 1).

Table 5. Global per-class accuracies, consistent-year land cover product (percent)												
IGBP Land Cover Class	Producer's Accuracy				User's Accuracy				Areal Proportions			
	Est.	Std. Err.	CI -	CI +	Est.	Std. Err.	CI -	CI +	Est.	Std. Err.	CI -	CI +
1. Evergreen Needleleaf	60.0	1.0	58.0	62.0	75.8	1.0	73.8	77.9	4.9	0.1	4.7	5.1
2. Evergreen Broadleaf	90.3	0.5	89.2	91.4	92.7	0.3	92.0	93.4	9.8	0.1	9.7	10.0
3. Deciduous Needleleaf	57.7	2.8	52.2	63.3	42.3	3.2	36.0	48.7	0.9	0.1	0.8	1.1
4. Deciduous Broadleaf	34.0	1.5	31.0	37.1	43.3	1.7	40.0	46.7	1.4	0.1	1.3	1.5
5. Mixed Forest	61.5	1.1	59.4	63.6	58.7	1.0	56.7	60.7	4.3	0.1	4.1	4.4
6. Closed Shrubland	14.2	1.1	12.1	16.3	46.2	3.0	40.3	52.2	1.9	0.1	1.7	2.1
7. Open Shrubland	85.0	0.6	83.7	86.3	46.8	1.0	44.8	48.7	9.6	0.2	9.2	9.9
8. Woody Savanna	51.6	1.4	48.8	54.4	37.5	1.2	35.1	39.8	4.2	0.1	4.0	4.5
9. Savanna	52.4	1.4	49.6	55.1	39.1	1.4	36.4	41.8	4.6	0.1	4.3	4.8
10. Grasslands	66.2	1.2	63.7	68.7	55.3	1.2	53.0	57.6	5.6	0.1	5.4	5.9
11. Permanent Wetlands	37.9	2.7	32.6	43.2	62.6	2.9	56.8	68.4	0.5	0.0	0.4	0.6
12. Cropland	58.1	0.6	56.8	59.4	87.4	0.4	86.5	88.3	14.0	0.2	13.7	14.3
14. Cropland/Nat Veg Mosaic	42.5	1.1	40.2	44.8	53.5	1.6	50.4	56.6	3.9	0.1	3.7	4.1
15. Snow and Ice	96.6	0.4	95.9	97.4	99.9	0.1	99.8	100	10.8	0.0	10.7	10.9
16. Barren/Sparse	74.8	0.7	73.4	76.2	92.8	0.4	92.0	93.6	14.9	0.1	14.6	15.2
17. Water	98.3	0.2	97.9	98.8	99.4	0.1	99.2	99.6	8.0	0.0	8.0	8.1

Producer's accuracies range from a low of 14.2 percent for closed shrubland, to 98.3 percent for water. The low value for closed shrubland occurs because many training pixels of this class are confused with open shrubland (Table 3). The next-lowest value is for deciduous broadleaf forest, which is typically confused with mixed forest. Note also the small number of pixels in this training class (Table 3). Like water, evergreen broadleaf forest and snow and ice also have high producer's accuracies.

User's accuracies are somewhat less variable, ranging from 39.1 percent for savanna to 99.9 percent for snow and ice. From the confusion table (Table 3), we note that savanna is often mapped as woody savanna or cropland. Other classes with low user's accuracies are deciduous needleleaf forest, deciduous broadleaf forest, closed shrubland, and open shrubland. These are typically confused with evergreen needleleaf and mixed forest; mixed forest and cropland-natural vegetation mosaic; open shrubland; and barren and cropland; respectively.

In general, producer's and user's accuracies are fairly similar for most classes. Where omission and commission errors are quite different, however, so are the two corresponsing accuracies. For example, closed shrubland shows a producer's accuracy of 14.2 percent and a user's accuracy of 46.2 percent. Here most of the errors are by omission rather than commission, and many closed shrubland pixels were classified into open shrubland. For closed shrubland, the reverse is true, with producer's accuracy at 85 percent and user's accuracy at 46.8 percent. Here the errors are mostly of commission. As noted above, a significant number of barren pixels are being classified into open shrubland.

While most estimated areal proportions shown in Table 5 are similar to those reported by pixel counts from the global map (Table 1), there are significant differences for open shrubland and cropland. Open shrubland is observed at 17.4 percent of pixels, while its actual estimated areal proportion is 9.6 percent. Here we see the errors of commission, documented by the low user's accuracy of 46 percent, pulling down the areal estimate. For cropland, the reverse is true. While 9.3 percent of pixels are classified as cropland, the estimator reports 14.0 percent cropland. In this case, it is the high omission error rate for cropland that boosts the areal estimate.

CONFIDENCE VALUES

As discussed in a preceding section, another statistic of use in evaluating a classification product is the confidence value, which is an output of the classifier that expresses the probability that the pixel being classified matches the training pixels input to the classifier.(Here values are expressed as percents.) Table 6 presents average confidence values by land cover classes, which range from 52.3 percent for permanent wetlands to 90 percent for barren lands. The average accuracy of all land cover types is 70.7 percent, while an average weighted by the estimators of global areal proportions (Table 5) is 78.3 percent.

Note that the confidence value for water is not available due to difficulties with the land-water mask that is overlain on the classification output. The area-weighted average uses a confidence value of 90 percent for water, based on its high producer's and user's accuracies.

Table 6. Global confidence values by land cover class (percent)

IGBP Land Cover Class	Average Confidence Value
1. Evergreen Needleleaf	68.3
2. Evergreen Broadleaf	89.3
3. Deciduous Needleleaf	66.7
4. Deciduous Broadleaf	65.9
5. Mixed Forest	65.4
6. Closed Shrubland	60.0
7. Open Shrubland	75.3
8. Woody Savanna	64.0
9. Savanna	67.8
10. Grasslands	70.6
11. Permanent Wetlands	52.3
12. Cropland	76.4
14. Cropland/Natural Veg	60.7
15. Snow and Ice	87.2
16. Barren	90.0
17. Water	(Not Available)
Average Value, All Classes	70.7
Area-Weighted Average	78.3

Table 7. Global confidence values by continental regions (percent).	
Region	**Average Confidence Value**
Global	76.3
Africa	79.4
Austr & Insular Asia	83.2
Eurasia	76.8
North America	71.9
South America	78.5

DISCUSSION

Global and Regional Accuracies

It is important to note that the accuracies reported in Tables 4 and 5 are valid only to the extent that the confusion table constitutes a random sampling of pixels and thereby captures the true variability of the map. However, in this case, the training samples cannot be considered a random sample.

Given the high cost of high-resolution imagery on which to delineate training sites, we selected many of our training sites using Landsat scenes obtained for other projects, both at Boston University and at other locations by other research groups. With scenes at hand from a particular continental region, we simply looked for "good training sites" within these scenes for the land cover types found within the region. Selection of sites also depended on the availability of good ancillary information. In addition to these "typical" sites, we also placed many sites in regions where either the classifier was having difficulty or we anticipated difficulty due to the complexity of the region. Often this required purchasing Landsat scenes to provide training sites in the problem areas. Thus, our training samples cannot be considered to be either random samples or even a representative sample of our consistent-year product.

Although the impact of this problem on our accuracy statistics is uncertain, it is likely that our training site selection has biased the results to the low side. Most classes will have large core areas in which they are well-developed and easily recognized, and a random sample can be expected to draw many pixels from such core regions. However, we tend to have relatively fewer training sites from these areas, since they are relatively homogeneous and provide little new information to the classifier. In fringe areas, where classes intergrade and transitions are frequent, we have proportionally more training sites, and in these regions, errors are likely to be more frequent. In addition, a small increment in accuracy may be inferred from the fact that the final classification uses all training sites, whereas the confusion matrix uses decision trees that omit ten percent of the sites.

The former estimate is also very consistent with the area-weighted confidence average of 78.3 percent, shown in Table 6 above. These values significantly exceed the error reported for a proper stratified random sample of the IGBP DISCover Land Cover product derived from AVHRR, which is 66.9 percent (Scepan, 1999).

Accuracies of Individual Land Cover Types

As reported above, the accuracy estimates for individual land cover types vary widely when taken from the confusion matrix. In many cases, this represents the inability of the classifier and the input data stream to differentiate consistently among intergrading types. For example, closed shrubland, open shrubland, grassland, savanna, and woody savanna can be taken as gradations of a land surface with varying amounts of tree, shrub, and grass cover. Distinctions among these types are difficult to make with coarse-resolution spectral-temporal data alone. Significant gradation also occurs between evergreen needleleaf, deciduous broadleaf, and mixed forest in North America, with deciduous needleleaf substituting for evergreen needleleaf forest in Eurasia. Another group of interrelated classes includes grassland, cropland, and cropland/natural vegetation mosaic.

In spite of these difficulties, some classes are recognized with high accuracy. These include evergreen broadleaf forest, snow and ice, barren/sparse, and water. Given their persistent and distinctive spectral signatures, it is not surprising that these classes are consistently classified correctly.

We should also note here that our training site database needs further improvement, including the addition of training sites in small classes. For example, the database includes only 15 training sites and 261 pixels for deciduous needleleaf forest. While this type is rather rare, as most of the Siberian larch forest is actually woody savanna by the IGBP definition, clearly more training sites are needed before we can have any faith in the estimated accuracies of this class. This is also the case with permanent wetlands, for which the database includes 13 sites and 289 pixels. In fact, the wide variance in training sites and pixels across classes precludes a formal analysis of individual errors, such as the probability that a pixel is actually grassland when it has been classified as cropland, or the probability that a pixel is actually evergreen needleleaf forest when it is actually mixed forest.

Given the inadequacies of the training site data, both from the viewpoint of its departure from a random sample and the fact that it significantly undersamples some classes, we think that the per-class confidence values, shown in Table 6, are likely to better represent the range of producer's and user's accuracies that are truly characteristic of our consistent-year land cover product. These vary from 60 percent for closed shrubland to 90 percent for barren/sparse, and conform well with our opinion of the classifier's ability and our map's accuracy. This is also about the range in accuracies observed for the IGBP DISCover dataset (Scepan, 1999).

REFERENCES CITED

Card, D. H, 1982, Using known map category marginal frequencies to improve estimates of thematic map accuracy, Photogramm. Engr. Remote Sens., vol. 48, pp. 431–439.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schnieder, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C., 2002, Global land cover from MODIS: Algorithms and early results, *Remote Sens. Environ.*, vol. 83, pp. 135-148.

McIver, D. K., and M. A. Friedl, 2002, Using prior probabilities in decision-tree classification of remotely sensed data, Remote Sens. Environ., vol. 81, pp. 253–261.

Scepan, J., 1999, Thematic validation of high-resolution global land-cover data sets, Photogramm. Engr. Remote Sens., vol. 65, pp. 1051–1060.